ESTIMATING INDICATORS OF LABOUR INPUT FROM ADMINISTRATIVE RECORDS HAVING COVERAGE AND MEASUREMENT ERRORS

Piero Demetrio Falorsi, Alessandro Pallara, Raffaella Succi, Italian National Institute of Statistics, and Aldo Russo, University of Roma Tre, Italy
Alessandro Pallara, Italian National Institute of Statistics, via Depretis 74/b, Rome, 00184, Italy
pallara@istat.it

ABSTRACT

Administrative data represent a valuable source for obtaining timely and detailed estimates of business population aggregates. A method is proposed herein for short-term estimation of two labor input indicators for subpopulation of firms. The estimates are obtained using a combination of cross-sectional estimates — obtained through a regression model on a subset of the sampled population containing data for the reference period — which borrow strength from the availability of current and past data on the population from Social Security Database. This will allow to obtain population estimates in each reference period accounting for a component which depends on information on business demography. Final estimates are then achieved adjusting to the target population totals obtained from the Italian Business Register, in order to correct for omissions and listing errors in the administrative source.

Key Words: **Administrative data, Short-term indicators, Small area estimation, Repeated surveys, Business registers**

1. INTRODUCTION

Recently, an increasing attention has been devoted to data obtained from administrative sources as a means for improving quality of survey data, especially for surveys of businesses, farms and institutions. There are some important advantages deriving from the use of administrative data: (*i*) the statistical information may be available at relatively little additional cost from several different sources; (*ii*) the timeliness is often better than can be achieved in complex surveys; (*iii*) information is obtained without additional loading on respondents; (*iv*) because of detailed coverage of administrative data they can be useful for obtaining estimates for small domains of statistical units.

This paper proposes a method for short-term estimation of two labor input indicators using data from Social Security Database (SSD) and from Italian Business Register (BR). The rationale for developing this methodology is originated by the need of the Italian national statistical office to meet the standard quality requirements in the *Regulation No. 1165/98* of the European Community concerning short-term business statistics. Information requested in the Regulation involves such a detailed disaggregation that it would be impossible to meet all the requirements through direct data collection.

SSD represents the most important administrative source in Italy for data on employment and wages. SSD contains: (*i*) current data on a subset of units that represents a *non random* sample of the population; (*ii*) information on the entire population with data referring to one year preceding the current period. Using these data for the purpose of constructing labor input indicators involves a number of problems, among others: elementary units in the database do not correspond to a definition of the statistical unit (*enterprise*) suitable for the analysis, the database do not match exactly the target population because of coverage and measurement errors. BR employment data, although having a better coverage of the target populations, cannot be utilized directly for the construction of the indicators, because of time delay in register updating. The estimates are obtained through a two-steps procedure: (*a*) *preliminary estimates* are yielded using only data from SSD, through a combination of cross-sectional estimates which borrow strength from the availability of current and past data; (*b*) *final estimates* are obtained adjusting preliminary estimates to the target population totals, modeling the association between historical data in the SSD and the BR through methods for the analysis of association between contingency tables.

The plan of the paper is as follows. Section 2 introduces the parameters of interest and describes the type of administrative sources used for the estimation. Then section 3 presents a detailed description of the estimation method. Finally, section 4 contains some concluding remarks related to the application of the proposed methodology.

2. PARAMETERS OF INTEREST AND SOURCES OF THE DATA

Let P_t be the population of enterprises of interest for current time interval t (month, e.g., with t=1,2,…) and $P_{t,w}$ (w=1,…,W) a given sub-population of P_t defined by the cross-classification of 60 industry types (first two digits of NACE rev. 1 nomenclature) with 20 geographical areas and a four-class categorical variable for employment size. The parameters of interest are the totals for the variables *employment* and *wage and salaries* for both the total population P_t and the sub-populations $P_{t,w}$. These parameters may be represented by the following expression

$$Y_{t,w} = \sum_{k \in P_{t,w}} y_{t,k} \qquad (t=1,2,\ldots;\ w=1,\ldots,W) \tag{1}$$

where $y_{t,k}$ denotes the value of the variable of interest y for the enterprise k in month t.

The method for estimating $Y_{t,w}$, described in the subsequent sections, is based on the use of BR and of three data sets of SSD. The data sets of SSD are: (*i*) the first data set, denoted hereafter as A_t, is a non random sample of P_t (roughly 100,000 enterprises) having information on employment and wages and salaries; (*ii*) the second, denoted as B_{t-12}, has data on the variables of interest for all the enterprises in the SSD in month t-12. Note that B_{t-12} may not give a completely adequate representation of P_{t-12} (the population of interest at month t-12) because of coverage problems depending on omission or late fulfilment for some enterprises of administrative duties in Social Security Regulations; (*iii*) the third data set, hereafter C_t, has information on all enterprises included in the register of Social Security in month t; for each unit, C_t contains, *inter alia*, data on geographical location, economic activity and employment at the time t^* (different for each enterprise) in which the enterprise first registered in SSD. Because of problems with keeping-up-to-date SSD (with entries to, removals from, transformation of units) C_t may have both coverage and measurement errors and, therefore, it may differ systematically from the target population P_t.

Indeed, BR has a better coverage of the target population although for a reference time t_0 preceding 24 or 36 months current time t. BR is the result of a complex set of record linkage and statistical procedures applied to data from business surveys and several administrative sources, including Social Security. BR, denoted hereafter as D_{t_0}, contains, for each enterprise, data on location, economic activity as well as number of employees which are updated yearly, with a reference time t_0.

3. ESTIMATION METHOD

As mentioned previously, the estimates are obtained through a two-steps procedure: (*i*) *preliminary estimates* are yielded using only data from SSD, through a combination of cross-sectional estimates which borrow strength from the availability of longitudinal information; (*ii*) *final estimates* are obtained adjusting preliminary estimates to the target population totals, modelling the association between SSD and BR in order to correct for non sampling errors in SSD.

3.1. Preliminary Estimates

Let us denote with: (*i*) $C_{t-12,t} = C_t \cap B_{t-12}$ the set of units that are common to B_{t-12} and C_t; (*ii*) $N_{t-12,t} = C_t - C_{t-12,t}$ the units that are included in C_t but are not included in B_{t-12} (enterprises entering into P_t during the time interval (t-12 , t)); (*iii*) $C_{(t-12,t),\tilde{w}}$ ($\tilde{w}$=1,…,$\tilde{W}$) a subset of $C_{t-12,t}$ ($C_{(t-12,t),\tilde{w}}$ may include, e. g., the enterprises of $C_{t-12,t}$ belonging to a given industry type) containing the enterprises that, according to the information recorded in C_t (economic activity and geographical areas) and in B_{t-12} (employment size), belong to a given sub-population, denoted also as *regression group* (note that the $\tilde{W}$ regression groups $C_{(t-12,t),\tilde{w}}$ define a complete partition of $C_{t-12,t}$); (*iv*) $N_{(t-12,t),\tilde{m}}$ ($\tilde{m}$=1,…,$\tilde{M}$) a *regression group* defined on the set $N_{t-12,t}$ (the $\tilde{M}$ regression groups $N_{(t-12,t),\tilde{m}}$ define a complete partition of $N_{t-12,t}$). Let us assume the following models hold:

$$y'_{t,k} = \beta_{t,\widetilde{w}}\ y'_{t-12,k} + \varepsilon_{t,k} \qquad (k \in C_{(t-12,\,t),\widetilde{w}}) \tag{2}$$

$$y'_{t,k} = \beta_{t,\widetilde{m}}\ x'_{t^*,k} + \nu_{t,k} \qquad (k \in N_{(t-12,\,t),\widetilde{m}}) \tag{3}$$

where $y'_{t,k}$ and $y'_{t-12,k}$ denote the values of the variable of interest for the k-th enterprise recorded in SSD databases (note that at current time t, $y'_{t,k}$ may be observed only for the enterprises included in A_t); $x'_{t^*,k}$ denotes the number of employees recorded in C_t in month t^* when the k-th enterprise first entered to the SSD; $\beta_{t,\widetilde{w}}$ and $\beta_{t,\widetilde{m}}$ are regression coefficients; $\varepsilon_{t,k}$ and $\nu_{t,k}$ are two random components with

$$\mathrm{E}(\varepsilon_{t,k}) = 0\,; \quad \mathrm{E}(\varepsilon_{t,k}, \varepsilon_{t,k}) = c_k\ \sigma^2_{\widetilde{w}}\,; \quad \mathrm{E}(\varepsilon_{t,k}, \varepsilon_{t,k'}) = 0 \qquad (k, k' \in C_{(t-12,\,t),\widetilde{w}}\ ; k \neq k')$$

$$\mathrm{E}(\nu_{t,k}) = 0\,; \quad \mathrm{E}(\nu_{t,k}, \nu_{t,k}) = d_k\ \sigma^2_{\widetilde{m}}\,; \quad \mathrm{E}(\nu_{k}, \nu_{k'}) = 0 \qquad (k, k' \in N_{(t-12,\,t),\widetilde{m}}\ ; k \neq k')$$

where c_k and d_k are known constants and $\sigma^2_{\widetilde{w}}$ and $\sigma^2_{\widetilde{m}}$ are the model variances. It is assumed that there is no serial correlation of the error term in equation (2), which includes a lagged endogenous variable in the right-hand side. With this assumption, estimation using either OLS or GLS yield unbiased and consistent estimates of the regression coefficients. An alternative to obtain consistent estimates of the β 's in presence of serially correlated errors is to use the technique of instrumental variables.

The regression groups should be as homogeneous as possible to yield (nearly) unbiased estimates at low levels of aggregation, while, conversely, the intersections sets $A_{(t-12,\,t),\widetilde{w}} = C_{(t-12,\,t),\widetilde{w}} \cap A_t$ and $A_{(t-12,\,t),\widetilde{m}} = N_{(t-12,\,t),\widetilde{m}} \cap A_t$ should have a sufficient sample size in order to provide good estimates of regression parameters $\beta_{t,\widetilde{w}}$ and $\beta_{t,\widetilde{m}}$. A good criterion for defining the regression groups is proposed in Hidiroglou, et al. (1995).

Let

$$\hat{y}'_{t,k} = \hat{\beta}_{t,\widetilde{w}}\ y'_{t-12,k} \qquad (k \in C_{(t-12,\,t),\widetilde{w}}) \tag{4}$$

$$\hat{y}'_{t,k} = \hat{\beta}_{t,\widetilde{m}}\ x'_{t^*,k} \qquad (k \in N_{(t-12,\,t),\widetilde{m}}) \tag{5}$$

be the predicted values of $y'_{t,k}$ in the models (2) and (3), respectively, where (cfr. Särndal et al., 1992, sez.6.4)

$$\hat{\beta}_{\widetilde{w}} = \left(\sum_{k \in A\,(t-12,t),\widetilde{w}} (y'_{t-12,k})^2 \Big/ c_k \right)^{-1} \sum_{k \in A\,(t-12,t)\,,\,\widetilde{w}} y'_{t-12,k}\ y'_{t,k} \Big/ c_k$$

$$\hat{\beta}_{\widetilde{m}} = \left(\sum_{k \in A\,(t-12,t),\widetilde{m}} (x_{t^*,k})^2 \Big/ d_k \right)^{-1} \sum_{k \in A\,(t-12,t),\widetilde{m}} x_{t^*,k}\ y'_{t,k} \Big/ d_k\ .$$

In order to yield preliminary estimates $\hat{Y}_{t,w}$ it is necessary to assign each unit in C_t to one of sub-populations $P_{t,w}$ (w=1,…,W) of interest. To this end: (i) the information on economic activity (two digits of NACE rev.1) and geographical location for each enterprise may be taken from C_t ; (ii) the information on class of employment may be obtained directly from the observed value, $y'_{t,k}$, if the k-th enterprise is included in A_t or, if the enterprise is not included in A_t , may be obtained by the predicted value using expressions (4) and (5) according to whether a unit is common to B_{t-12} and C_t or it is entering to P_t during the time interval (t-12 , t). It is then possible to associate to each enterprise in C_t an indicator variable $\delta_{t,k}(w)$ that is equal to 1 if the k-th enterprise belongs to the sub-population $P_{t,w}$ and equal to 0 otherwise. Using a predictive approach, the preliminary estimates $\hat{Y}_{t,w}$ are given by

$$\hat{Y}_{t,w} = \sum_{k\in A_t} y'_{t,k}\ \delta_{t,k}(w) + \sum_{\widetilde{w}=1}^{\widetilde{W}} \sum_{k\in \overline{A}_{(t-12,t),\widetilde{w}}} \hat{y}'_{t,k}\ \widetilde{\gamma}_{t,\widetilde{w}}\ \delta_{t,k}(w) + \sum_{\widetilde{m}=1}^{\widetilde{M}} \sum_{k\in \overline{A}_{(t-12,t),\widetilde{m}}} \hat{y}'_{t,k}\ \widetilde{\gamma}_{t,\widetilde{m}}\ \delta_{t,k}(w) \qquad (6)$$

where:

$$\overline{A}_{(t-12,t),\widetilde{w}} = C_{(t-12,t),\widetilde{w}} - A_{(t-12,t),\widetilde{w}} \quad ; \quad \overline{A}_{(t-12,t),\widetilde{m}} = N_{(t-12,t),\widetilde{m}} - A_{(t-12,t),\widetilde{m}}\ .$$

In (6) $\widetilde{\gamma}_{t,\widetilde{w}}$ represents the estimate of the probability $\gamma_{t,\widetilde{w}}$ that an enterprise included in $\overline{A}_{(t-12,t),\widetilde{w}}$ is still alive and active. The need for introducing these probabilities derive from the circumstance that the set $C_{(t-12,t),\widetilde{w}}$, from which the set $\overline{A}_{(t-12,t),\widetilde{w}}$ is defined, may include some enterprises that no longer belong to P_t; $\widetilde{\gamma}_{t,\widetilde{m}}$ has a similar meaning with respect to $N_{(t-12,t),\widetilde{m}}$. In order to estimate these probabilities we use the information in the data sets A_{t-12}, C_{t-12}, B_{t-12} and B_{t-24} (the content of data sets A_{t-12}, C_{t-12} and B_{t-24} has an obvious relation with data sets A_t, C_t and B_{t-12} introduced in the previous section).The procedure is articulated in the following steps:

(a) definition of the sets $C_{(t-24,t-12),\widetilde{w}}$ ($\widetilde{w}=1,...,\widetilde{W}$) and $N_{(t-24,t-12),\widetilde{m}}$ using information recorded in the data sets C_{t-12} and B_{t-24} and the same rules as those described above for defining the sets $C_{(t-12,t),\widetilde{w}}$ and $N_{(t-12,t),\widetilde{m}}$;

(b) attribution of the value of the indicator variable $\delta_{t-12,k}(\widetilde{w})$ to each of the enterprises in C_{t-12}. This may be done in the following way: (*b1*) the information on economic activity and geographical location may be taken from C_{t-12}; (*b2*) the information on employment class may be obtained directly from the observed value, $y'_{t-12,k}$, if the *k*–th enterprise is included in B_{t-12} or, if the enterprise is not included in B_{t-12}, by the value predicted using expressions similar to (4) or (5) above, computed for a different time interval;

(c) the probabilities $\gamma_{t,\widetilde{w}}$ and $\gamma_{t,\widetilde{m}}$ may be estimated by

$$\widetilde{\gamma}_{t,\widetilde{w}} = \sum_{k\in B_{t-12}\cap C_{(t-24,t-12),\widetilde{w}}} \delta_{t-12,k}(\widetilde{w}) \Big/ \sum_{k\in C_{(t-24,t-12),\widetilde{w}}} \delta_{t-12,k}(\widetilde{w}) \qquad (7)$$

$$\widetilde{\gamma}_{t,\widetilde{m}} = \sum_{k\in B_{t-12}\cap N_{(t-24,t-12),\widetilde{m}}} \delta_{t-12,k}(\widetilde{m}) \Big/ \sum_{k\in N_{(t-24,t-12),\widetilde{m}}} \delta_{t-12,k}(\widetilde{m}). \qquad (8)$$

The estimates $\widetilde{\gamma}_{t,\widetilde{w}}$ and $\widetilde{\gamma}_{t,\widetilde{m}}$ represent nearly unbiased estimates of $\gamma_{t,\widetilde{w}}$ and $\gamma_{t,\widetilde{m}}$ only if these probabilities have low variability over the time span between one estimate and the following. An alternative solution with repeated surveys, very interesting but computationally very complex, consists in formulating the problem of the estimate of $\gamma_{t,\widetilde{w}}$ and $\gamma_{t,\widetilde{m}}$ as solution of a dynamic state-space model (Tam 1987; Binder, et al. 1993).

3.2. Final Estimates

As mentioned in the previous section, the preliminary estimates are affected by: (*i*) *classification errors*, i.e. an enterprise may belong to a sub-population different from that assigned using information recorded in SSD; (*ii*) *measurement errors,* i.e. the value of a variable of interest, $y'_{t,k}$, recorded in SSD may differ from the true value, $y_{t,k}$; (*iii*) *under-coverage errors*, namely SSD may not include some enterprises of P_t that have not complied their administrative duties of Social Security Regulations.

The second step of the estimation procedure is meant for adjusting preliminary estimates, $\hat{Y}_{t,w}$, for the bias introduced by these errors. Final estimates are then computed considering three multiplicative factors, each one aiming to correct for any specific source of error. In order to describe this step, it is useful to introduce two matrices obtained by linking D_{t_0} and B_{t_0} (B_{t_0} is analogous to B_{t-12}, but with reference to month t_0):

$$\boldsymbol{\Omega}_{t_0} = (Y_{t_0,(w,l)}) \ (w=1,\ldots,W;\ l=1,\ldots,W+1) \quad ; \quad \boldsymbol{\Psi}_{t_0} = (Y'_{t_0,(w,l)}) \ (w=1,\ldots,W+1;\ l=1,\ldots,W)$$

where: $Y_{t_0,(w,l)}$ denotes the total of the variable of interest, obtained using information in D_{t_0}, for the enterprises classified in the sub-population $P_{t_0,w}$ in D_{t_0}, and classified in the sub-population $P_{t_0,l}$ according to B_{t_0}; $Y'_{t_0,(w,l)}$ indicates the total for the same enterprises calculated with information in B_{t_0}. The (w,W+1)-th cell of $\boldsymbol{\Omega}_{t_0}$ is the total for units not included in B_{t_0} but included in D_{t_0} and herein classified in the sub-population $P_{t_0,w}$. The (W+1, l)-th cell of $\boldsymbol{\Psi}_{t_0}$ contains the total (calculated with information collected in B_{t_0}) for units not included in D_{t_0} and classified in B_{t_0} in sub-population $P_{t_0,l}$.

Final estimates are obtained by means of a three steps procedure, assuming throughout that associative structure between $\boldsymbol{\Omega}_{t_0}$ and $\boldsymbol{\Psi}_{t_0}$ observed at time t_0 is still valid at current time t.

In the first step an estimate of $\boldsymbol{\Psi}_t$ is obtained ($\boldsymbol{\Psi}_t$ is analogous to $\boldsymbol{\Psi}_{t_0}$ but referred to time t). Using the matrix $\boldsymbol{\Psi}_{t_0}$ and the preliminary estimates $\hat{Y}_{t,l}$ (l=1,…,W) calculated from (6), the cell total $\hat{Y}'_{t,(w,l)}$ of $\boldsymbol{\Psi}_t$ is estimated as

$$\hat{Y}'_{t,(w,l)} = \hat{Y}_{t,l}\ {}_1\lambda_{(w,l)} \quad (w=1,\ldots,W+1;\ l=1,\ldots,W) \tag{9}$$

where

$${}_1\lambda_{(w,l)} = Y'_{t_0,(w,l)} \Big/ \sum_{w=1}^{W+1} Y'_{t_0,(w,l)} . \tag{10}$$

This step is intended for correcting *classification errors* in SSD.

Second step is then intended to correct for *measurement errors*. In this step the first W rows of the matrix $\boldsymbol{\Omega}_t$ ($\boldsymbol{\Omega}_t$ is analogous to $\boldsymbol{\Omega}_{t_0}$ but referred to time t) are calculated as

$$\hat{Y}_{t,(w,l)} = \hat{Y}'_{t,(w,l)}\ {}_2\lambda_{(w,l)} \qquad (w=1,\ldots,W;\ l=1,\ldots,W) \tag{11}$$

in which

$${}_2\lambda_{(w,l)} = Y_{t_0,(w,l)} \Big/ Y'_{t_0,(w,l)} .$$

(12)

First two steps do not allow to correct for *under-coverage* represented by the W+1-th column of $\boldsymbol{\Omega}_t$. In order to solve this problem a third set of correcting factors, ${}_3\lambda_{(w,l)}$ (w=1,…,W;l=1,…,W), is introduced. These factors are obtained as solution of the following constrained minimum problem

$$\begin{cases} \displaystyle\sum_{w=1}^{W}\sum_{l=1}^{W} L(Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}\ {}_3\lambda_{(w,l)}, Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)})\, c_{w,l} = \min \\ \displaystyle\sum_{l=1}^{W} Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}\ {}_3\lambda_{(w,l)} = Y_{t_0,w} \quad w=1,\ldots,W \end{cases} \tag{13}$$

in which $L(Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}\ {}_3\lambda_{(w,l)}, Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)})$ is a distance function between $Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}\ {}_3\lambda_{(w,l)}$ and $Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}$; $c_{w,l}$ is a known weight, depending from the size of the (w,l)-th cell. Then, from (13), cell totals $Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}\ {}_3\lambda_{(w,l)}$ are computed which are as close as possible to totals ($Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}$) obtained through the previous steps, thus correcting for the under-coverage in SSD. Choosing a specific distance function introduces in (13) a particular association models among the values $Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}\ {}_3\lambda_{(w,l)}$ and $Y'_{t_0,(w,l)}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}$ (cfr. Little and Wu, 1991).

Having defined the correcting factors ${}_1\lambda_{(w,l)}$, ${}_2\lambda_{(w,l)}$ and ${}_3\lambda_{(w,l)}$ (w=1,…,W;l=1,…,W), final estimates are then yielded as:

$$\tilde{Y}_{t,w} = \sum_{l=1}^{W} \hat{Y}_{t,l}\ {}_1\lambda_{(w,l)}\ {}_2\lambda_{(w,l)}\ {}_3\lambda_{(w,l)} \tag{14}$$

4. DISCUSSION AND SOME CONCLUDING REMARKS

In this paper we have illustrated a methodology for obtaining estimates of labour input indicators using data from administrative sources. Short term statistics on employment and wages have severe coverage problems in Italy, in that the only sample surveys currently executed cover only very large businesses (>500 employees) and specific groups of economic activities. Using administrative data has in this case the advantage that definitions are coherent with the requirements of the European Community Regulation and it is therefore possible to obtain necessary information without additional burden on enterprises. However, as mentioned at the beginning, there are shortcomings concerning the characteristics of administrative data, which limit their use for statistical purposes. These limitations deserve careful examination, especially from the point of view of a data collection agency whose responsibility involves the quality of the measurement process for data dissemination. Using data from the SSD encompasses the following issues: (*i*) the subset of data relative to the current month is not a random sample, selected according to a specific design; (*ii*) concepts and measured variables may follow a different concept of data definition which can not be controlled by the statistician.

The methodology developed in this paper tries to overcome some of these shortcomings, proposing *ad hoc* solutions, taking into account the characteristics of the particular administrative source. However, there are some open problems with the estimation method herein proposed, including: (*i*) the selection of the regression groups in (2) and (3); (*ii*) the definition of a particular distance function and a choice of the cell weight $c_{w,l}$ in (13); (*iii*) the estimation of mean squared error of the final estimate $\tilde{Y}_{t,w}$ using either linearization techniques or methods based on replications.

REFERENCES

Binder D.A., S.R. Bleuer, and J.P. Dick (1993), "Time Series Methods Applied to Survey Data," *Bulletin of the International Statistical Institute*, 49(1), pp. 327-344.

Hidiroglou, M.A., M. Latouche, B. Armstrong, and M.Gossen (1995), "Improving Survey Information Using Administrative Records: the Case of the Canadian Employment Survey," *Proceedings of the Annual Research Conference, U.S Bureau of the Census*, pp. 171-197.

Little R.J.A., and M. M. Wu (1991), "Models for Contingency Tables With Known Margins When Target and Sampled Population Differ," *Journal of the American Statistical Association*, 86, pp. 87-95.

Särndal C.E., B. Swensson, and J. Wretman (1992), *Model Assisted Survey Sampling*, New York: Springer.

Tam, S.M. (1987), "Analysis of Repeated Surveys Using a Dynamic Linear Model," *International Statistical Review*, 55, pp. 63-73.